Syntactic variation as a measure of probabilistic indigenization in global varieties of English

Jason Grafmiller, Benedikt Heller, Melanie Röthlisberger, Benedikt Szmrecsanyi

University of Leuven
Quantitative Lexicology and Variational Linguistics

Outline

Introduction

Data & methods

Results

Conclusion

Current project

- "Exploring probabilistic grammar(s) in varieties of English around the world" (5-year project, 2013–2018; PI: Benedikt Szmrecsanyi)
- synthesize disjoint lines of scholarship—research on World Englishes & probabilistic theories of grammar—into unified project with a coherent empirical and theoretical focus
- main goal: understand the plasticity of grammatical knowledge among language users from diverse regional and cultural backgrounds

The World Englishes paradigm

- study of the wide range of postcolonial varieties (e.g. Hong Kong E), inner circle varieties (e.g. British E), shift varieties (e.g. Irish E), ...
- topics: scope, limits, parameters of variation
  - extent to which structural make-up of varieties of E can be predicted by communicative needs of colonizers/colonized (e.g. Kachru 1992; Schneider 2007; Mesthrie 2008)
- empirical focus on the variable presence/absence of features, or in usage frequencies of features

Theoretical framework

adopt a variation-centered, usage- and experience-based probabilistic grammar framework (e.g. Bod et al. 2003; Bresnan et al. 2007; Bybee 2010)

1. syntactic variation and change is **subtle, gradient & probabilistic** (Labov 1982; Bresnan and Hay 2008; Wolk et al. 2013)
2. linguistic knowledge includes **knowledge of probabilities**, and speakers have powerful predictive capacities (Gahl and Garnsey 2004; Gahl and Yu 2006)

Research questions

- scope and limits of variation: To what extent do the varieties of English we study here share a stable probabilistic grammar?
- dialect typology: Does variety type (e.g. L1 vs. L2, inner vs. outer circle) predict similarity among certain varieties?
- variation phenomena: Do the alternations under study differ in terms of their sensitivity to variety effects? Are certain predictors more open to variation than others?

Outline

Introduction

Data & methods

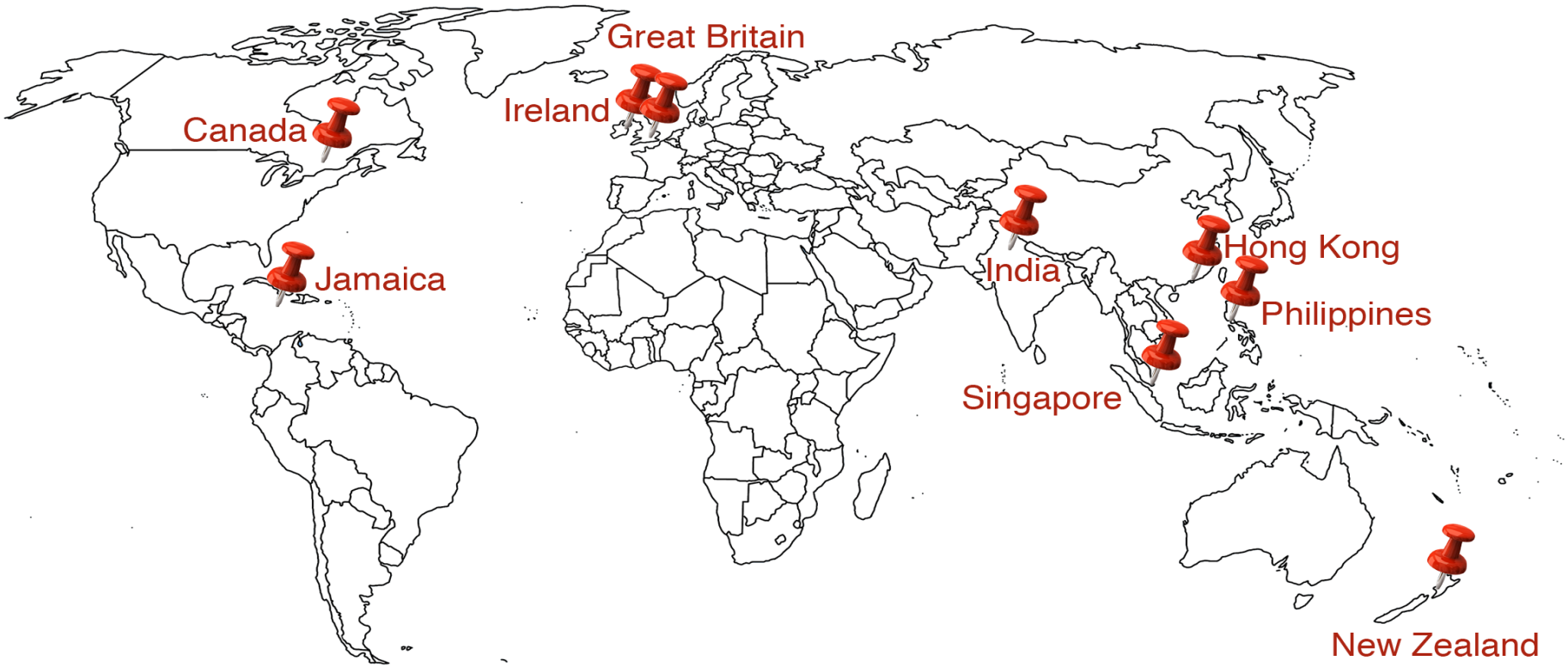
Results

Conclusion

Methodological sketch

1. tap into the **International Corpus of English (ICE)** to explore **3 syntactic alternations** across **9 global varieties**
2. create richly annotated datasets to study the interplay of the factors constraining syntactic choices
   - look for significant differences among individual varieties and across language groups
3. conduct supplementary rating-task experiments
   - do participants' judgments align with corpus model predictions?

Corpus data



- each ICE component contains ∼1.5 mil. words sampling 15 spoken and written genres
- currently adding web-based language from GloWbE

Syntactic phenomena

3 *very* well-known syntactic alternations

1. genitive alternation
2. dative alternation
3. particle placement alternation

Genitive alternation ($N = 10{,}592$)

(1) a. My niece's new chainsaw is orange. [*s*-genitive]

b. The new chainsaw of my niece is orange. [*of*-genitive]

- examine ordering of possessor and possessum
- variable context: excluded partitive genitives, pronominal genitives, and indefinite possessums

Dative alternation ($N = 8{,}549$)

(2) a. I gave my niece a chainsaw for Christmas. [ditransitive]

b. I gave a chainsaw to my niece for Christmas. [prepositional]

- examine ordering of recipient and theme
- variable context:
  - started from list of 80-some dative verbs (*give*, *take*, *tell*, etc.)
  - manually filtered benefactives, passives, and sentences with extracted or elliptical arguments

Particle placement ($N = 8{,}072$)

(3) a. My niece picked up her new chainsaw. ['joined' V-Prt-NP]

b. My niece picked her new chainsaw up. ['split' V-NP-Prt]

- examine ordering of particle and direct object
- variable context:
  - transitive particle verbs involving one of 10 most frequent particles:
    around, away, back, down, in, off, out, over, on, up
  - manually filtered passives, sentences with extracted or elliptical objects, prepositional arguments, etc.

Internal predictors (see Rosenbach 2014; Bresnan et al. 2007; Gries 2003)

Genitives	Datives	Part. Placement
possr./possm. animate?	rec./theme animate?	idiomaticity
possr./possm. length	rec./theme length (ratio)	D.O. length
possr. given?	rec./theme given?	D.O. given?
final sibilant?	rec./theme pronominal?	directional PP?
type-token ratio	rec./theme definite?	D.O. definite?
possr. frequency	rec./theme complex?	D.O. frequency
	rec./theme concrete?	D.O. concrete?
	rec. local? (1-2 person)	Prt. probability (ΔP)

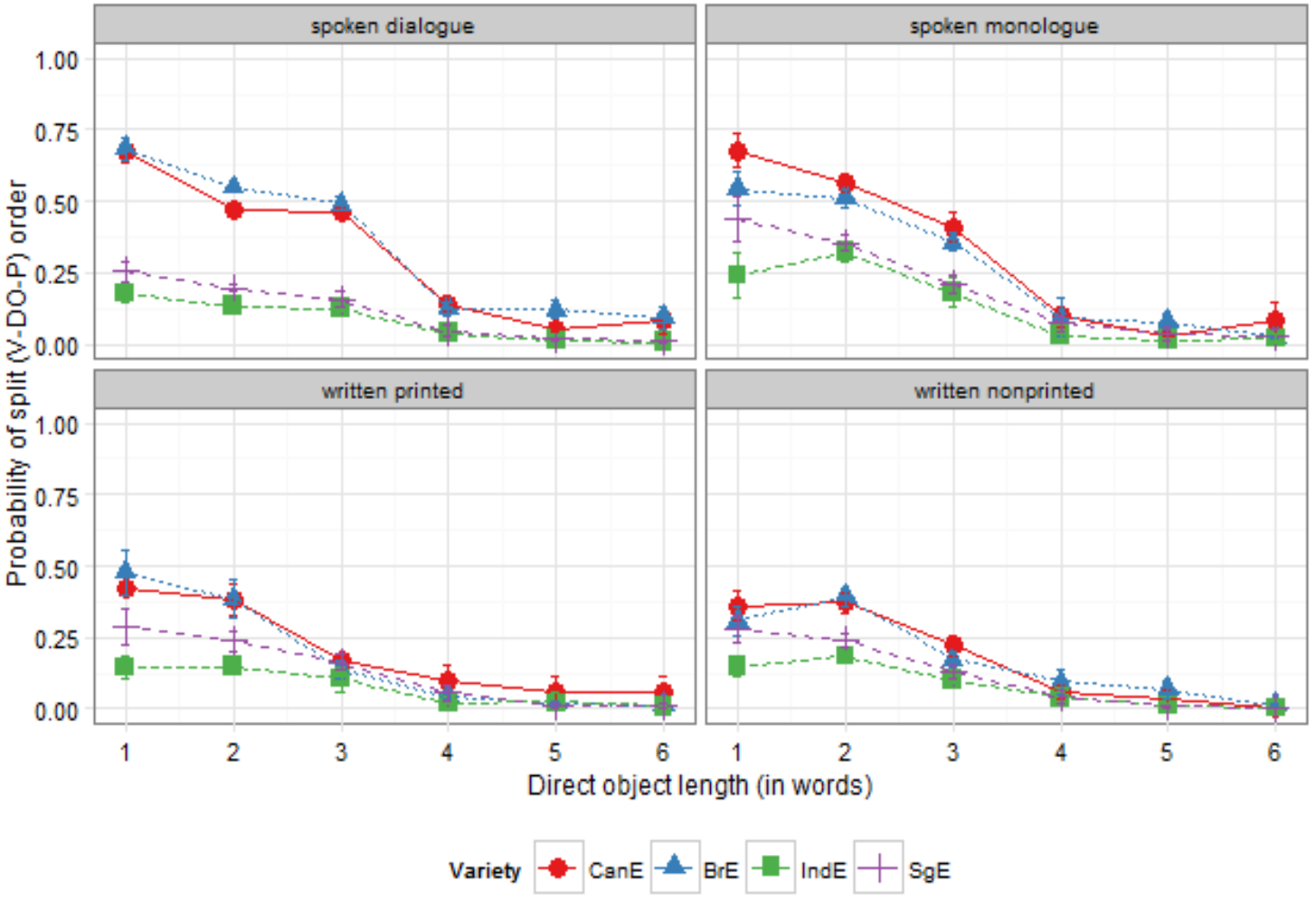
- treat other factors—e.g. verb, genre, speaker, text—as 'random effects' (for now)

Analysis: some old and new techniques

- statistically model the the influence of various constraints on a binary outcome
  - Mixed-effects logistic regression
  - Conditional random forests (Tagliamonte and Baayen 2012)
  - Memory-based learning (Daelemans and van den Bosch 2009; Theijssen et al. 2013)
  - ...
- examine the variability of internal predictors across varieties, genres, regions, registers
- relatively scant sociolinguistic metadata available for speakers

Pilot study

- influence of direct object length on particle placement is weaker in IndE and SgE than BrE and CanE



Outline

Introduction

Data & methods

Results

Conclusion

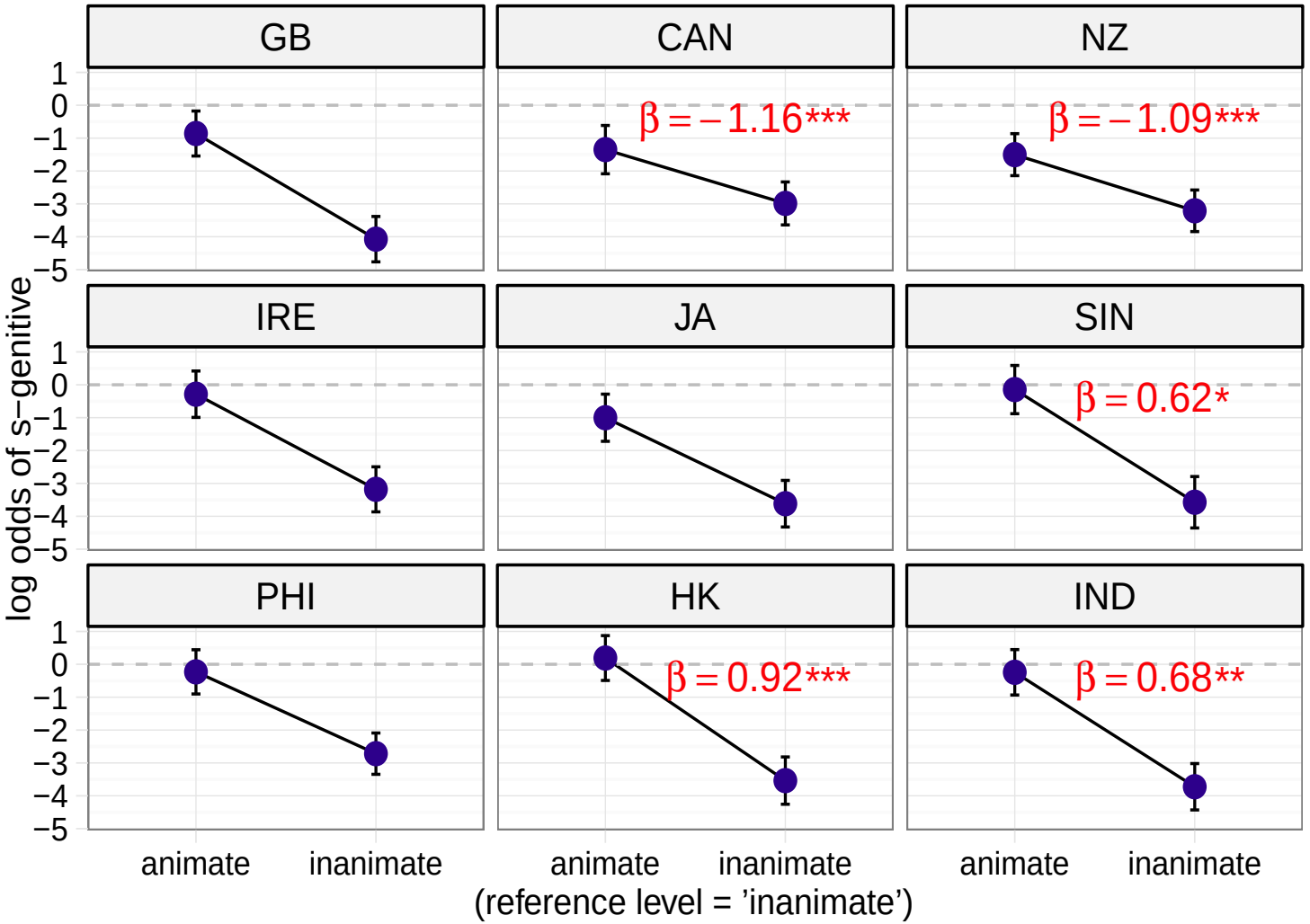
Findings so far

- alternations are fairly stable across varieties
  - clear, variety-independent, patterns among predictors ("main effects")
- the effect directions of factors are stable across varieties of English but some differences with regard to *effect size*
- no clear generalizations across individual varieties

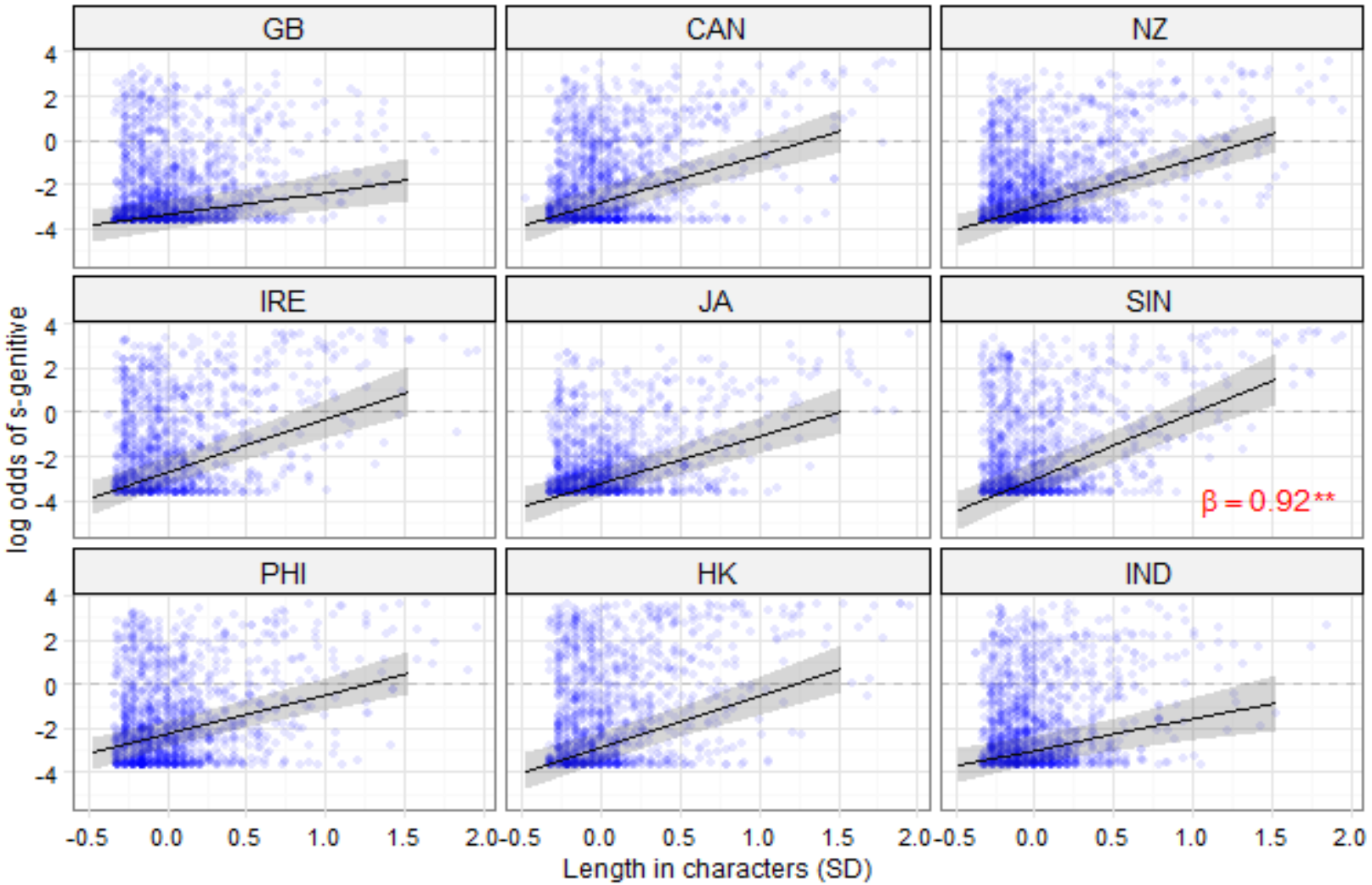
Probabilistic differences in the genitive alternation

- **animate possessors** favor *s*-genitive:
  - weaker influence in NZE and CanE
  - stronger influence in IndE, HKE, and SgE
- **final sibilant** favors *of*-genitive:
  - weaker influence in IndE and NZE
  - Stronger influence in CanE
- **longer possessums** favor *s*-genitive:
  - weaker influence in BrE
  - stronger influence in SgE

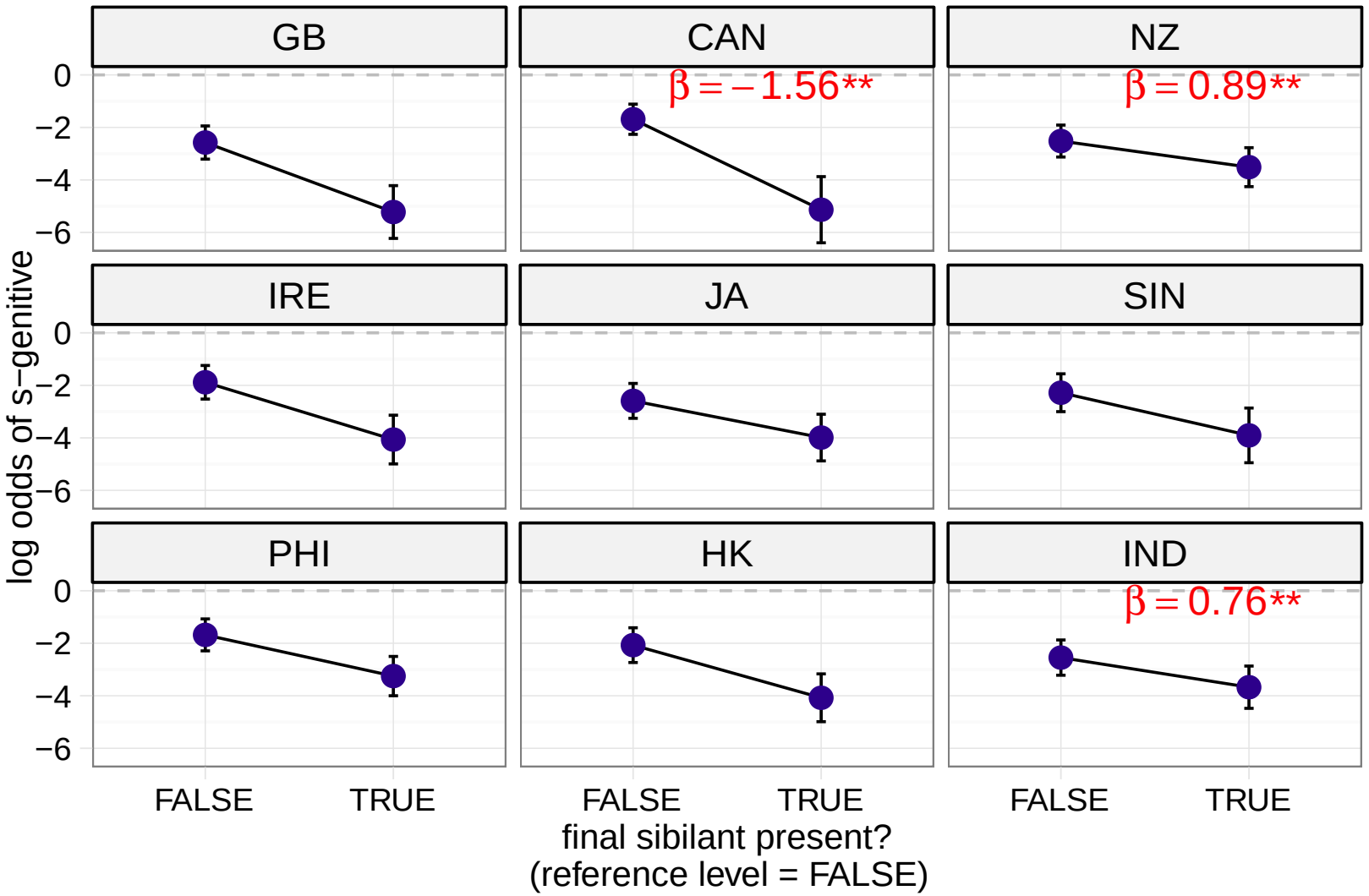
genitive alternation: Possessor animacy



genitive alternation: Possessum length



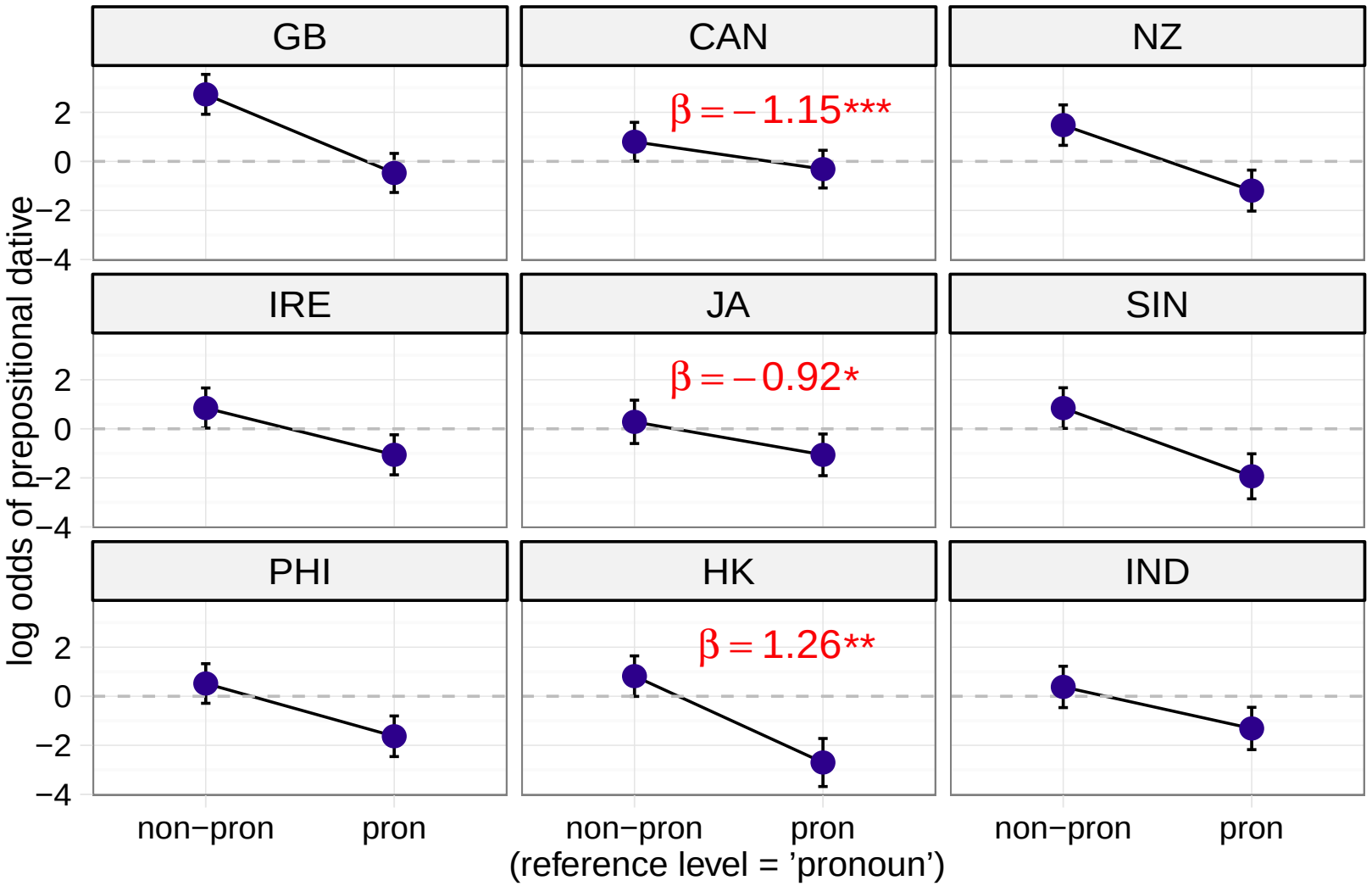
genitive alternation: Final sibilant



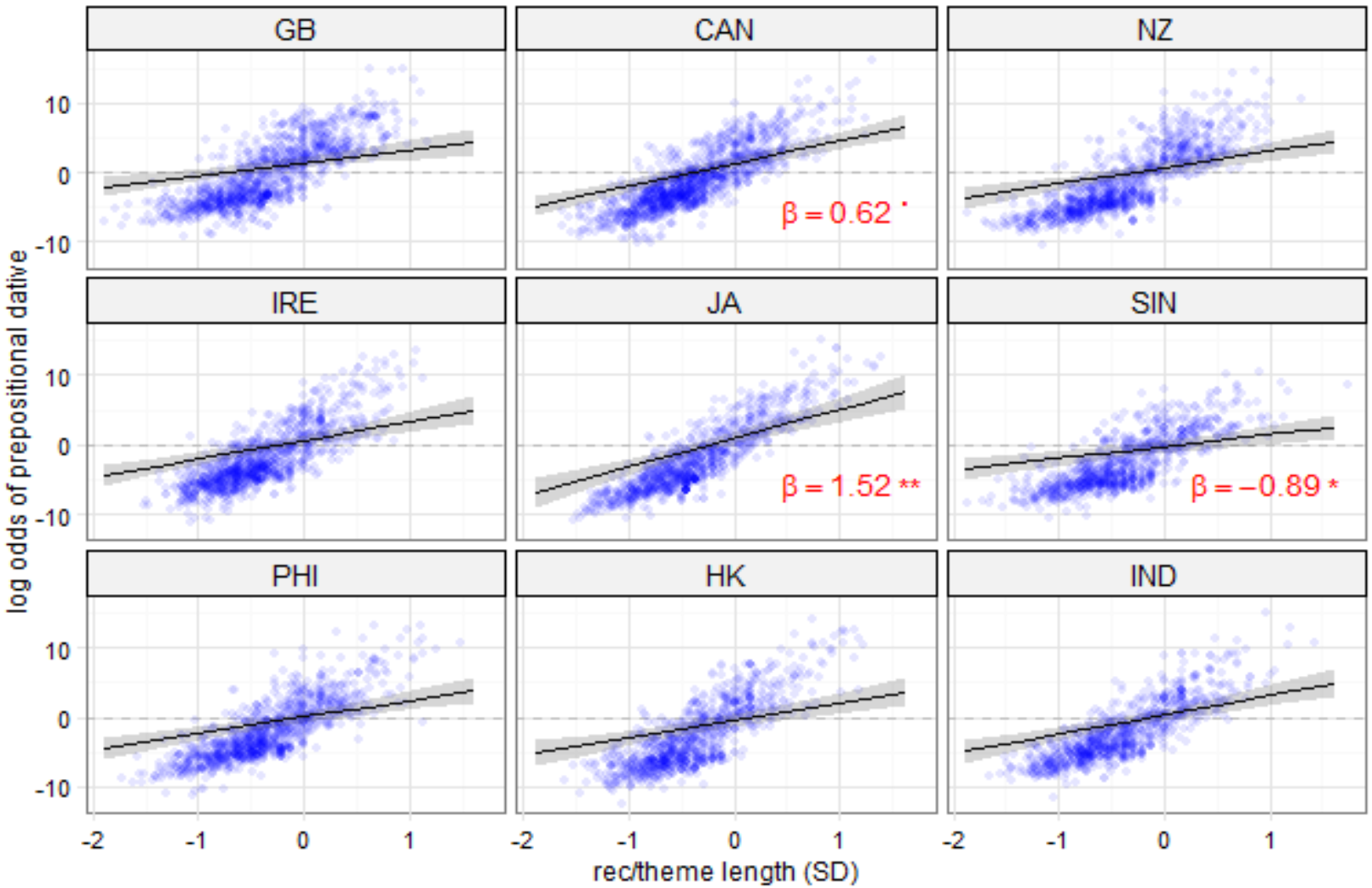
Probabilistic differences in the dative alternation

- **pronominal recipients** favor ditransitive:
  - stronger influence in HKE
  - weaker influence in JamE and CanE
- **longer recipients** favor prepositional dative:
  - stronger influence in SgE
  - weaker influence in JamE
- **concrete themes** favor prepositional dative:
  - weaker influence in HKE

dative alternation: Recipient pronominality



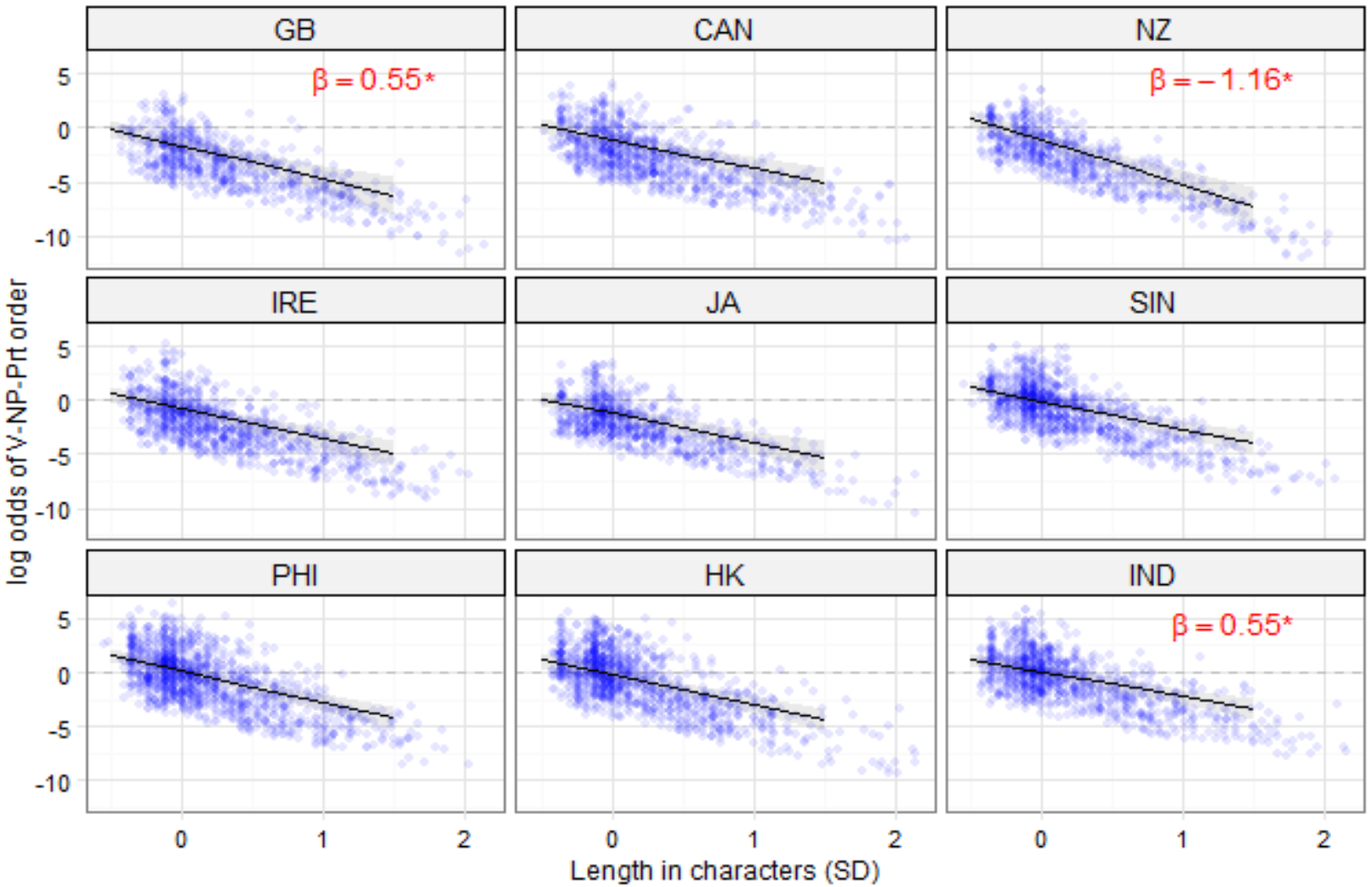
dative alternation: Recipient - theme length



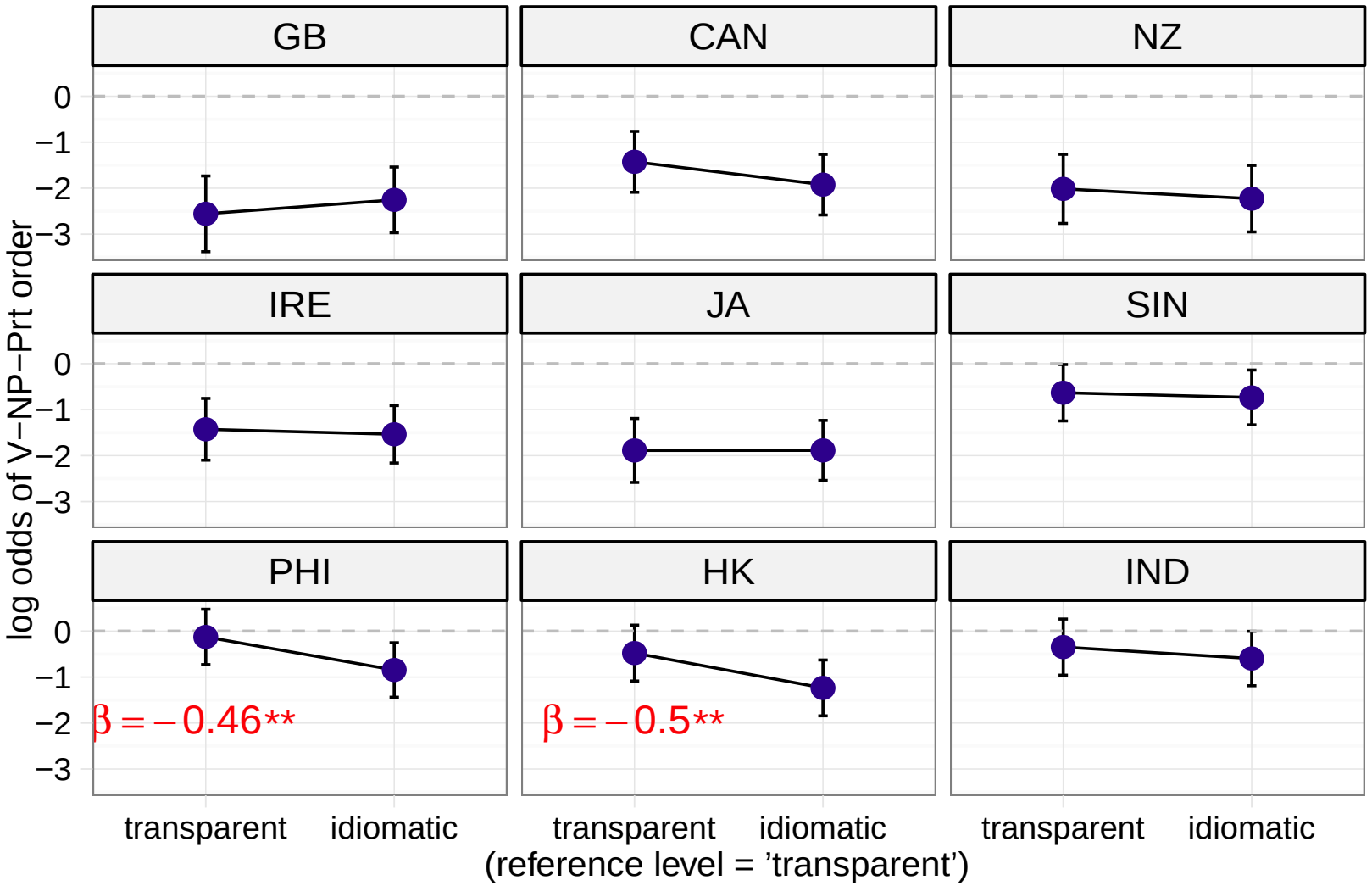
Probabilistic differences in particle placement

- **longer direct objects** favor V-Prt-NP:
  - weaker influence in NZE
  - stronger influence in IndE
- **Idiomatic verbs** favor V-Prt-NP:
  - stronger influence in HKE and PhiE
- **concrete objects** favor V-NP-Prt:
  - stronger influence in NZE and PhiE
- **directional PP** favors V-NP-Prt:
  - stronger influence in JamE
- stronger **verb-particle collocation** (ΔP)) favors V-Prt-NP:
  - weaker effect in JamE

particle placement: Direct object length



particle placement: Idiomaticity



Outline

Introduction

Data & methods

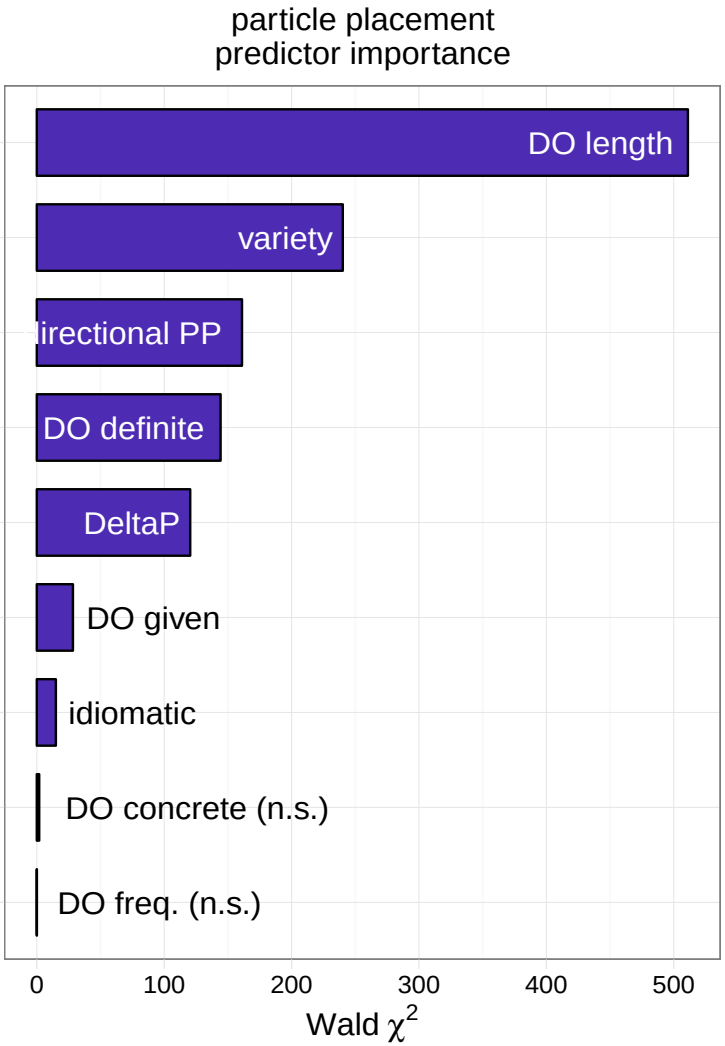
Results

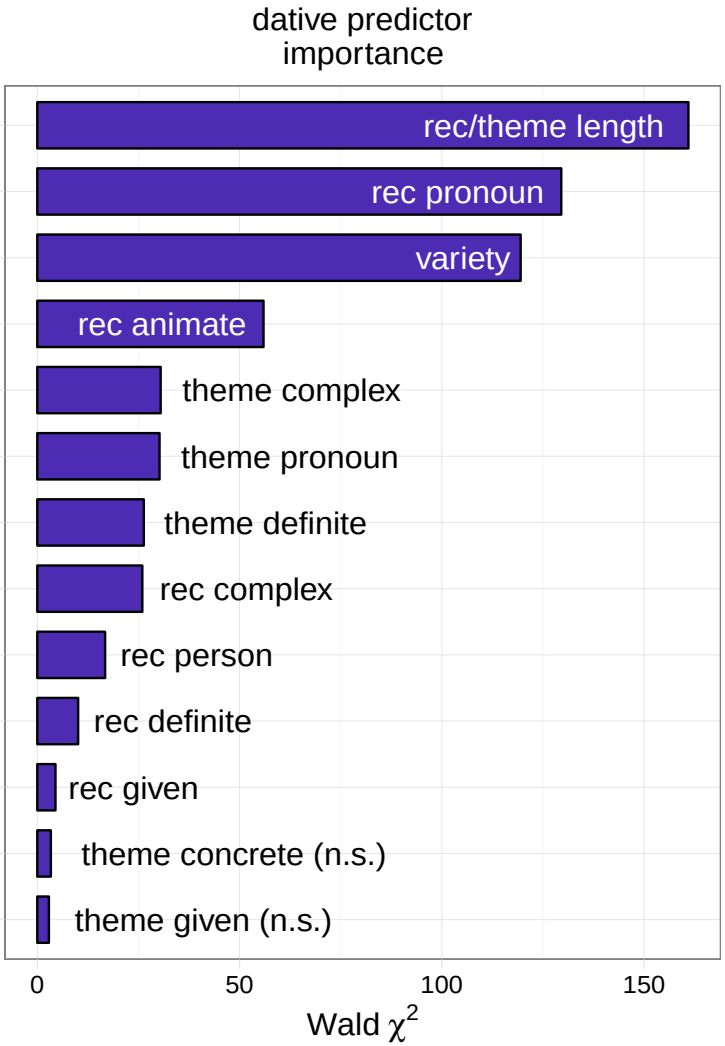
Conclusion

Findings thus far

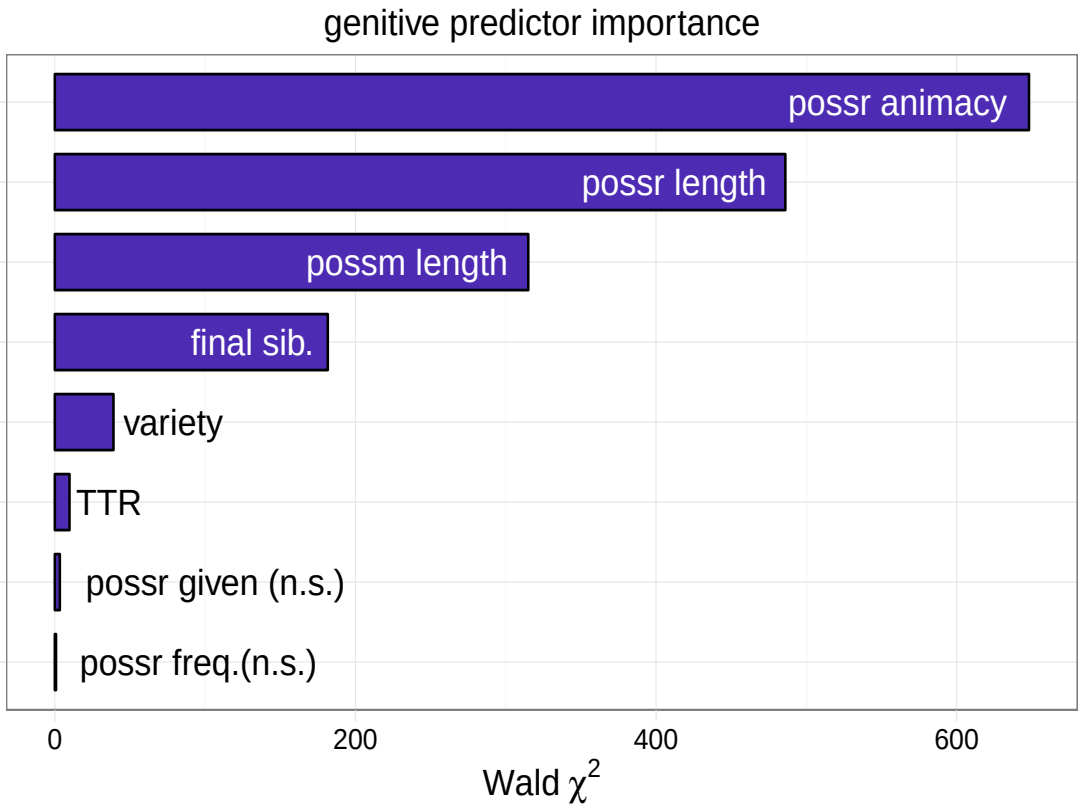
- grammars are fairly stable across varieties
  - reliable variety-independent patterns among predictors ("main effects")
- the effect directions of factors are stable across varieties of English but some differences with regard to *effect size*
- L2 varieties do not consistently differ from L1 varieties
- influence of variety on specific alternations differs a lot

Explanatory power of 'variety'





Explanatory power of 'variety'



Alternations are not equally sensitive to variety effects

- vary in amenability to "**probabilistic indigenization**"

"the process whereby stochastic patterns of internal linguistic variation are reshaped by shifting usage frequencies in speakers of post-colonial varieties. To the extent that patterns of variation in a new variety A [. . .] can be shown to differ from those of the mother variety, we can say that the new pattern represents a novel, if gradient, development in the grammar of A."
(Szmrecsanyi et al. in press)

Alternation variability

- more amenable: particle placement, dative alternation(?)
- less amenable: genitive alternation
- hypotheses:
  - lexico-syntactic innovation: the more entrenched a given syntactic alternation is w.r.t. specific lexical items the more likely it is to exhibit cross-varietal indigenization effects (Grafmiller & Röthlisberger 2015)
  - L2 acquisition effects: alternations that are relatively difficult for L2 learners will exhibit more cross-varietal variability

What's next?

- correlations with specific models of postcolonial English development, e.g. Schneider's (2003, 2007) Dynamic Model?
- variability across registers and genres?
- comparison with learner English?
- substrate influence?
- do off-line ratings correlation with corpus probabilities?

Thank You!

This presentation is based upon work supported by an Odysseus grant of the Research Foundation Flanders (FWO) (grant no. G.0C59.13N).

contact info:

jason.grafmiller@kuleuven.be

http://wwwling.arts.kuleuven.be/qlvl/ProbGrammarEnglish.html

Selected References I

Bod, R., J. Hay, and S. Jannedy (Eds.) (2003). Probabilistic Linguistics. Cambridge, MA: MIT Press.

Bresnan, J., A. Cueni, T. Nikitina, and B. Harald (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, and J. Zwarts (Eds.), Cognitive Foundations of Interpretation, pp. 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J. and J. Hay (2008). Gradient Grammar: An Effect of Animacy on the Syntax of give in New Zealand and American English. Lingua 118, 245–259.

Bybee, J. (2010). Language, Usage and Cognition. Cambridge: Cambridge University Press.

Gahl, S. and S. Garnsey (2006). Knowledge of grammar includes knowledge of syntactic probabilities. Language 82(2), 405–410.

Grafmiller, J. and M. Röthlisberger (2015, October). Syntactic alternations, schematization, and collostructional diversity in world Englishes. In The 21st Conference of the International Association for World Englishes, Istanbul.

Gries, S. T. (2003). Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement. New York: Continuum Press.

Kachru, B. B. (Ed.) (1992). The Other tongue: English across cultures (2nd ed ed.). English in the global context. Urbana: University of Illinois Press.

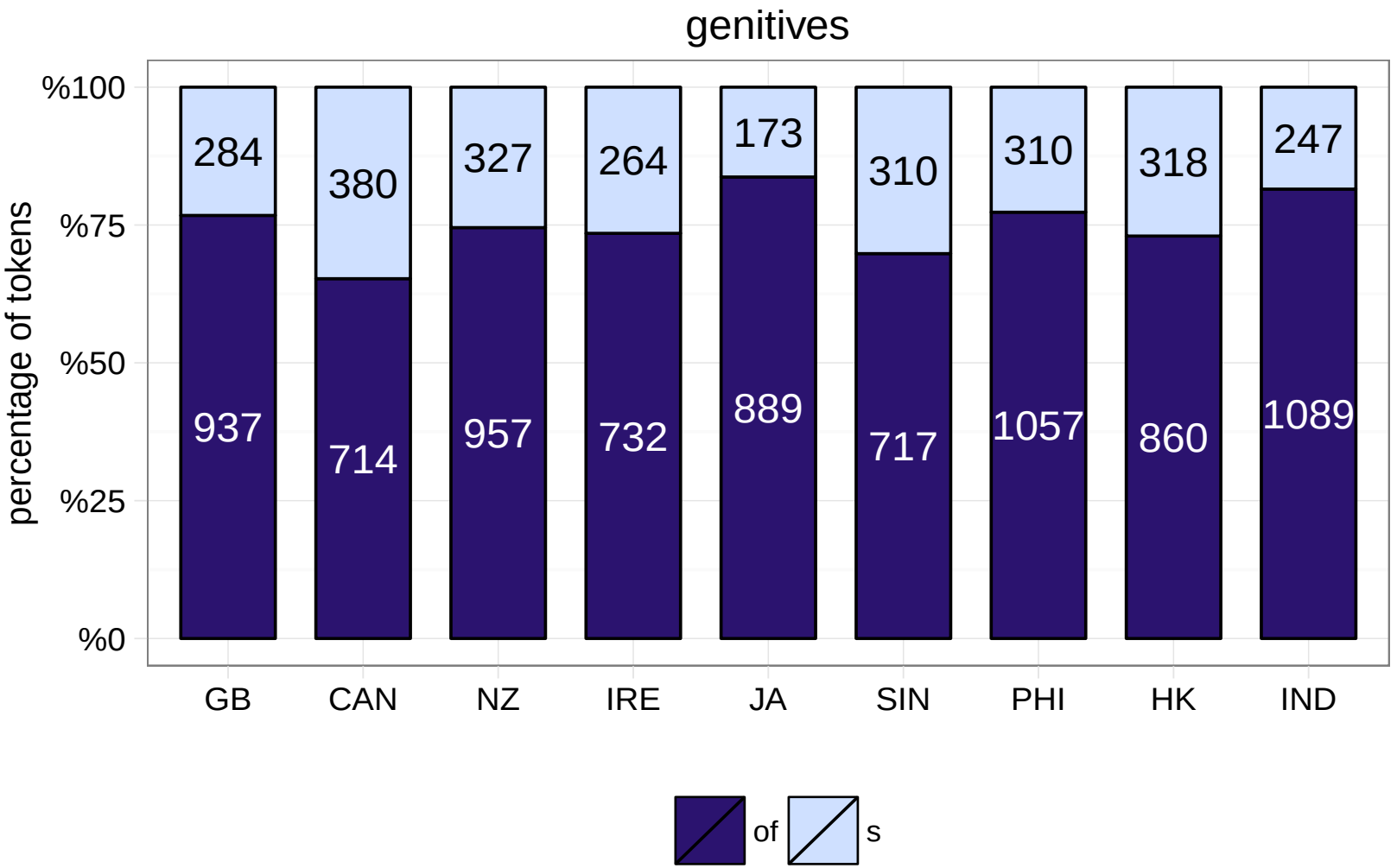
Rosenbach, A. (2014). English genitive variation – the state of the art. English Language and Linguistics 18(2), 215–262.

Schneider, E. (2003). The dynamics of New Englishes: from identity construction to dialect birth. Language 79, 281–298.

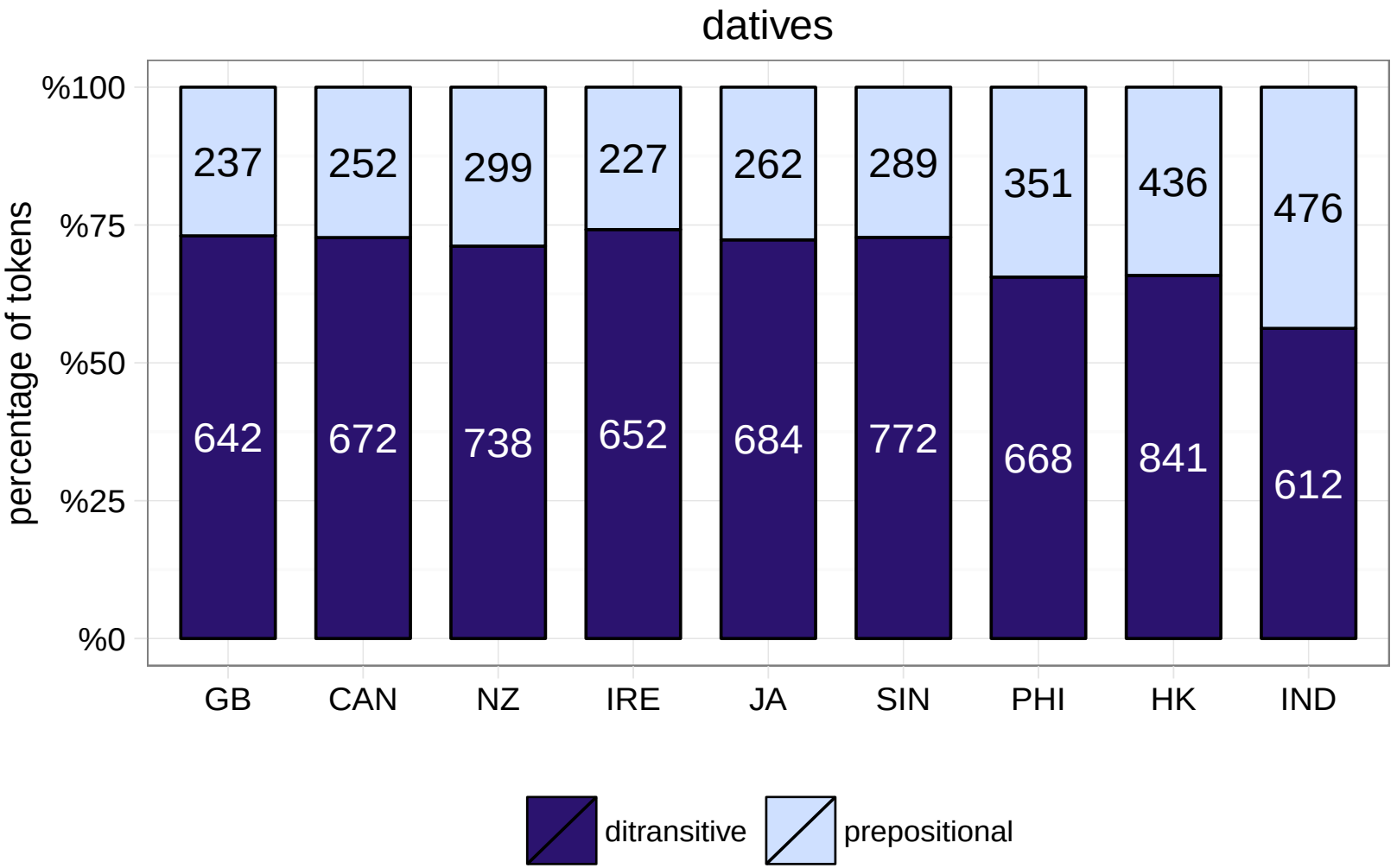
Schneider, E. (2007). Postcolonial English: Varieties Around the World. New York: Cambridge University Press.

Szmrecsanyi, B., J. Grafmiller, B. Heller,, and M. Röthlisberger (In press). Around the world in three alternations: modeling syntactic variation in global varieties of English. English World-Wide 37(2).

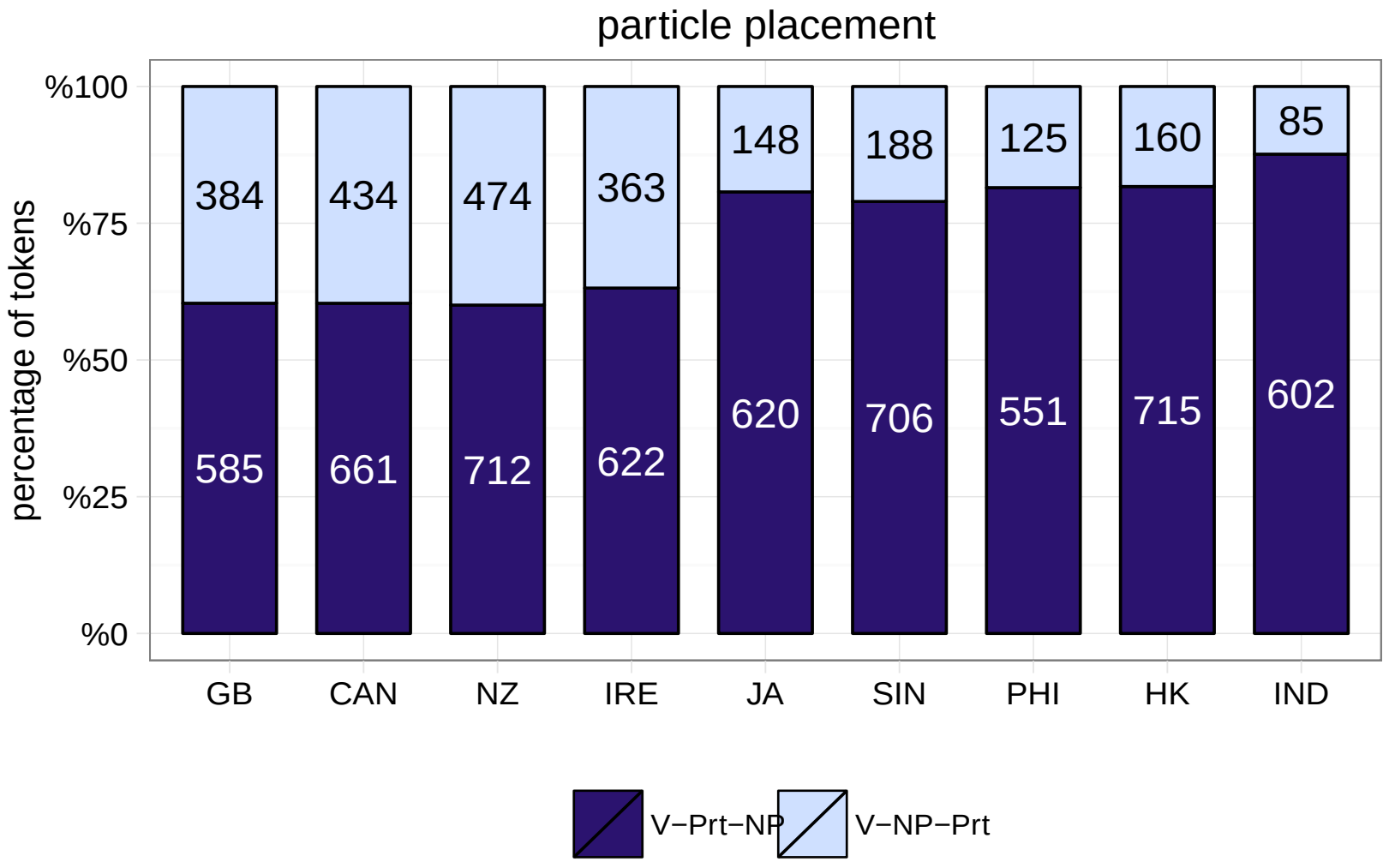
Genitive variants across varieties



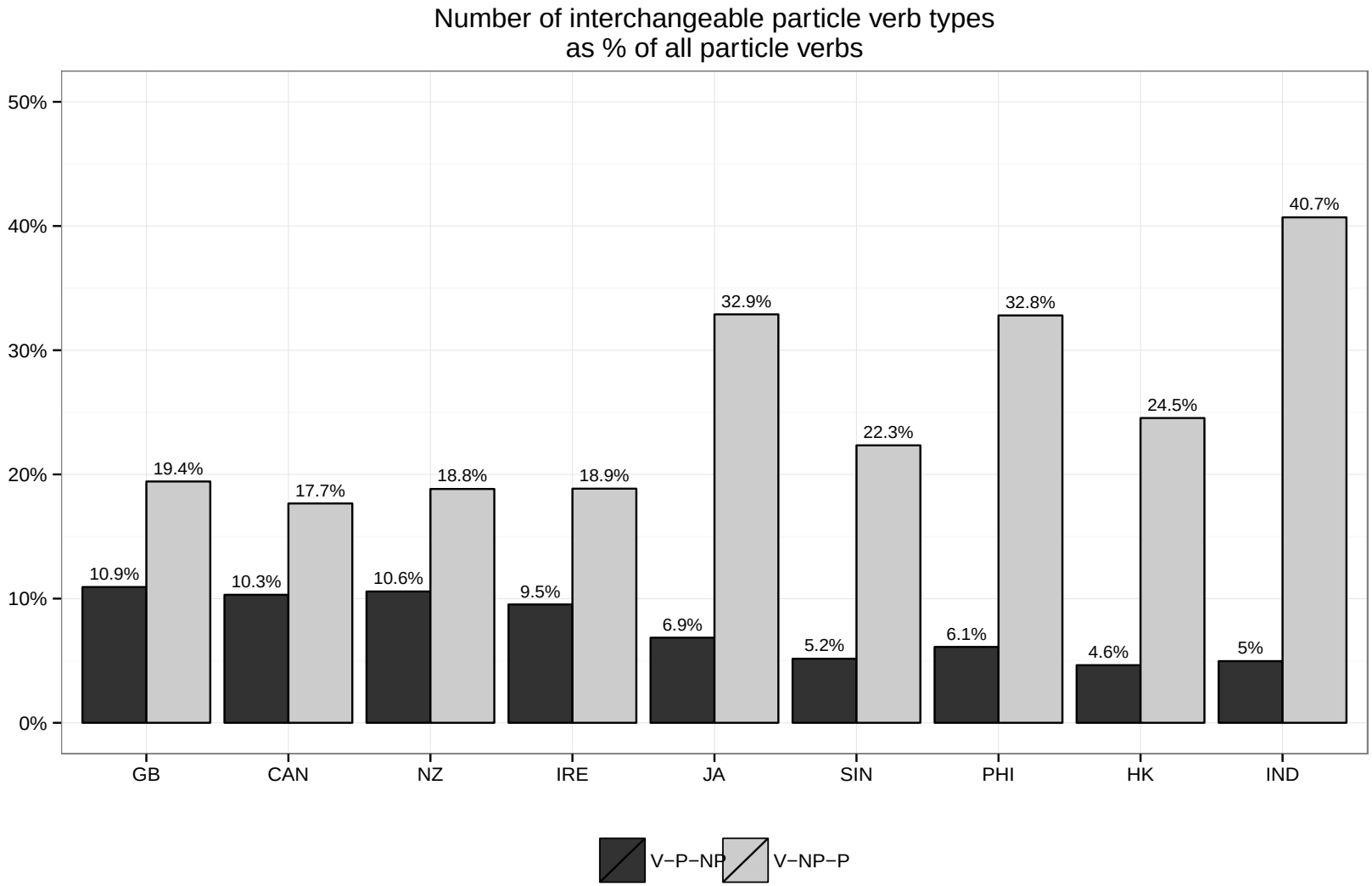
Dative variants across varieties



Particle placement across varieties



particle verbs: "allostructional" asymmetry across varieties



dative verbs: "allostructional" asymmetry across varieties